

Predictive model for the quantitative analysis of human skin using photothermal radiometry and diffuse reflectance spectroscopy

NINA VERDEL,^{1,2,5}  JOVAN TANEVSKI,³  SAŠO DŽEROSKI,^{3,4} 
AND BORIS MAJARON^{1,2,6} 

¹Department of Complex Matter, Jožef Stefan Institute, Jamova cesta 39, 1000 Ljubljana, Slovenia

²Faculty of Mathematics and Physics, University of Ljubljana, Jadranska 19, 1000 Ljubljana, Slovenia

³Department of Knowledge Technologies, Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia

⁴Jožef Stefan International Postgraduate School, Jamova cesta 39, 1000 Ljubljana, Slovenia

⁵nina.verdel@ijs.si

⁶boris.majaron@ijs.si

Abstract: We have recently introduced a novel methodology for the noninvasive analysis of the structure and composition of human skin *in vivo*. The approach combines pulsed photothermal radiometry (PPTR), involving time-resolved measurements of mid-infrared emission after irradiation with a millisecond light pulse, and diffuse reflectance spectroscopy (DRS) in the visible part of the spectrum. Simultaneous fitting of both data sets with respective predictions from a numerical model of light transport in human skin enables the assessment of the contents of skin chromophores (melanin, oxy-, and deoxy-hemoglobin), as well as scattering properties and thicknesses of the epidermis and dermis. However, the involved iterative optimization of 14 skin model parameters using a numerical forward model (*i.e.*, inverse Monte Carlo - IMC) is computationally very expensive. In order to overcome this drawback, we have constructed a very fast predictive model (PM) based on machine learning. The PM involves random forests, trained on ~9,000 examples computed using our forward MC model. We show that the performance of such a PM is very satisfying, both in objective testing using cross-validation and in direct comparisons with the IMC procedure. We also present a hybrid approach (HA), which combines the speed of the PM with versatility of the IMC procedure. Compared with the latter, the HA improves both the accuracy and robustness of the inverse analysis, while significantly reducing the computation times.

© 2020 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

We have recently introduced an innovative methodology for noninvasive assessment of structure and composition of human skin, based on combined pulsed photothermal radiometry (PPTR) and diffuse reflectance spectroscopy (DRS) [1–3]. PPTR involves measurements of transient dynamics in mid-infrared (IR) emission from the sample surface after exposure to a short light pulse [3–6]. This allows, *e.g.*, assessment of optical or thermal properties in homogenous samples, but not a unique physiological characterization of human skin, which contains multiple absorbing substances (melanin, oxy- and deoxy-hemoglobin, etc.) [5].

DRS in visible and near-IR part of the spectrum is a popular experimental technique in biomedical optics. However, it is well known that even for a semi-infinite homogeneous medium, the absorption and scattering coefficients can not be determined independently from a single DRS measurement without introducing additional assumptions (such as spectral constraints). Moreover, quantitative assessment of the chromophore contents in multi-layered organs requires either prior knowledge about their depth distribution at the investigated site (*e.g.*, thicknesses of the characteristic skin layers) [1,7,8], or spatially or temporally resolved measurements of DRS.

Combining PPTR and DRS overcomes their individual weaknesses by exploiting the respective strengths [1,3,9]. Such an approach allows simultaneous assessment of the contents of specific chromophores, thicknesses of the epidermis and dermis, as well as their scattering properties. We achieve this by fitting measured PPTR signals and DRS spectra with the corresponding predictions from a numerical model of light-tissue interaction (*i.e.*, inverse MC approach - IMC). To this end, we apply a four-layer optical model of skin, accounting for the epidermis, papillary and reticular dermis, and subcutaneous adipose layer [2,3,9].

However, the involved iterative multi-dimensional optimization using the MC forward model is computationally very expensive. Moreover, each optimization task must be repeated several times in order to control the inevitable numerical noise, facilitate escape from local minima, and assess the robustness of the IMC analysis. Despite massive parallelization (using a high-performance graphics card) assessment of 14 free parameters from each radiometric transient and DRS spectrum can take several hours, which seriously hinders practical exploitation of this approach.

In order to overcome this drawback we present here a computationally very efficient predictive model (PM) constructed by using machine learning. We cast the problem of learning the PM in the frame of multi-target regression and solve it using random forests, a robust and flexible approach offering good performance in various application areas [10]. Specifically, our PM learns the mapping from the joint space of PPTR signals and DRS spectra to the space of skin parameters, based on ~9,000 examples computed using the forward MC model.

We demonstrate that the performance of such PM is very satisfying, both in objective testing using cross-validation and in direct comparisons with the IMC approach when analyzing real (experimental) data. In short, the PM bypasses the demanding IMC optimization procedure and allows quick assessment of multiple properties of human skin from PPTR signals and DRS spectra measured *in vivo*.

We also present and test a hybrid approach (HA), which combines the unparalleled speed of the PM with versatility of the IMC procedure. Specifically, the HA uses the PM predictions to initialize the iterative IMC optimization. This yields results with lower residual norms and reduced variances compared with the customary IMC, while at the same time significantly reducing the computation times.

2. Methodology

2.1. Experimental protocol

The study involved healthy volunteers with fair skin (Fitzpatrick types I–II) and 25–65 years of age, and was approved by the Medical Ethics Committee of the Republic of Slovenia. Selected test sites on the subjects' arms were shaved and the dehydrated superficial layer (stratum corneum) was removed by tape stripping. This should ensure unimpaired heat diffusion all the way to the skin surface, as assumed in our mathematical model of the PPTR procedure. The site was then cleaned with medical-grade ethanol, rehydrated using physiological solution, and left to dry.

DRS in visible spectral range ($\lambda = 400\text{--}650\text{ nm}$) was measured using an integrating sphere (IS) with an internal light source and sample opening diameter of 10.3 mm (ISP-REF by Ocean Optics, Dunedin, FL). Spectral response of the spectrometer (USB4000, Ocean Optics) was calibrated using a Spectralon white standard. 100 subsequently acquired spectra (at integration time of 15 ms) were averaged to control the signal-to-noise ratio. The so-called single-beam substitution error, which is significant in such small IS, was removed by numerical pre-processing based on prior analysis [11], thus yielding artifact-free DRS values.

For PPTR measurements, the same test site was irradiated with individual 1 ms pulses at $\lambda = 532\text{ nm}$ from a medical-grade laser (DualisVP by Fotona d.o.o., Ljubljana, Slovenia). Such green light is absorbed by both epidermal melanin and hemoglobin in dermal vasculature. At the effective spot size of ~5 mm, the central radiant exposure was 0.20–0.30 J/cm².

The subsequent Mid-IR emission from the tissue surface was recorded with a fast mid-IR camera (SC7500, FLIR Systems; $\lambda = 3.5\text{--}5.1\ \mu\text{m}$) at a rate of 1000 frames per second. PPTR signals were obtained by lateral averaging of the radiometric data over a sub-window corresponding to an area of $\sim 1.5 \times 1.5\ \text{mm}^2$ on the skin surface and subtracting the pre-pulse baseline values [4,5]. Using the manufacturer provided calibration system (HypocalTM) the signal amplitudes were converted to radiometric temperature values.

The PPTR signals were recorded for at least 1.5 s and DRS was acquired for approx. 2 s. Both measurements were also repeated 3–5 times and the data was averaged before further processing. Our results thus represent skin property values averaged over physiological oscillations faster than 1–2 s, such as beating of the heart and breathing. Especially the former is known to induce considerable modulations of the blood content oxygenation, etc. If uncontrolled, these could adversely affect any quantitative analysis through the combined effects on the chromophore contents, tissue scattering, and penetration depth of the probing light [12].

2.2. Optical model of human skin

Based on our previous experience we apply a model of human skin with four optically homogenous layers, representing the epidermis, papillary and reticular dermis, and subcutis (see Fig. 1) [3,9,13]. Unlike the thicknesses of epidermis and entire dermis (d_{epi} and d_{der}), that of papillary dermis is fixed to $100\ \mu\text{m}$ to reduce the number of model parameters.

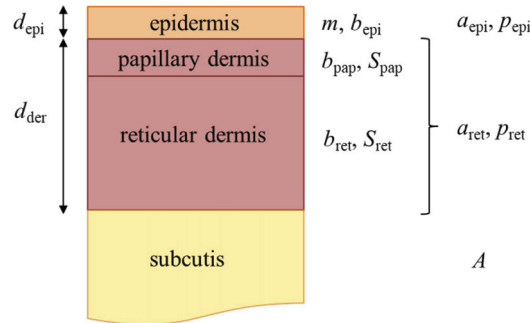


Fig. 1. Schematic presentation of our four-layer skin model and its free parameters (see the text for explanations).

In short, the absorption spectra of the epidermis and dermis are based on the widely accepted approach by Jacques [14], which combines a baseline spectrum with contributions due to specific contents of melanin in the epidermis (with volume fraction m) and blood in both dermal layers (b_{pap} and b_{ret} , respectively).

The absorption coefficient of blood is calculated as a linear combination of the values for oxygenated and deoxygenated whole blood [15], according to the oxygen saturation levels in the papillary and reticular dermis (S_{pap} and S_{ret}). We also account for optical screening in larger blood vessels by using the customary correction factor and typical vessel size distribution [16,17]. In addition, we allow for a small amount of blood in the epidermal layer (b_{epi}), to account for the undulation of the epidermal-dermal junction (dermal papillae).

The absorption spectrum of the subcutaneous adipose tissue is adopted from Simpson *et al.* [18].

In addition, we also optimize the scattering properties of the epidermis and dermis, according to the customary ansatz:

$$\mu'_s(\lambda) = a \left(\frac{\lambda}{500\ \text{nm}} \right)^{-p}. \quad (1)$$

This enables a significantly improved match with experimental data and more realistic results compared with the approach where $\mu_s'(\lambda)$ is fixed to average data from literature [3,9], in agreement with common knowledge that scattering properties of skin can vary between different subjects and anatomical sites within the same subject [14]. The respective parameters a and p are adjusted independently for epidermis and dermis.

Scattering properties of the subcutaneous adipose tissue are also adjusted individually, according to

$$\mu_{s, \text{sub}}'(\lambda) = A \left[16.43 \text{ cm}^{-1} + 303.8 \text{ cm}^{-1} \exp\left(-\frac{\lambda}{180.3 \text{ nm}}\right) \right], \quad (2)$$

where inter- and intra-personal variations are accounted for by the varying amplitude A . This function matches well the two scattering spectra reported by Salomatina *et al.* [19] at values of $A = 0.64$ and 1.5 , respectively [8].

The refractive index is set to $n = 1.45$ for the epidermis, 1.37 for dermis, and 1.34 for subcutis [20].

2.3. Modeling of light-tissue interaction

Light transport and energy deposition in skin during our measurements are simulated using the weighted-photon Monte Carlo (MC) technique [21]. More specifically, we apply the GPU-parallelized version of the original multi-layer MC code, derived by Alerstam *et al.* [22]. By considering also thermal properties of the involved tissues, the laser-induced energy deposition profiles can be converted into the corresponding temperature profiles. From these, the resulting PPTR signals can be computed as [4–6]

$$\Delta S(t) = \int_{z=0}^{\infty} K(z, t) \Delta T(z, 0), \quad (3)$$

where the Kernel function $K(z, t)$ accounts for spectrally dependent attenuation of the IR emission from subsurface tissue layers [6].

We use the same MCML model also to derive the DRS spectra for the same skin structure and composition. In doing so, we also account for the finite sample opening of our IS (see Sect. 2.1).

All MC simulations involve following the quasi-random tortuous paths and energy deposition of 10^7 energy packets (“photons”) at each considered wavelength.

2.4. Assessment of skin properties by inverse Monte Carlo (IMC)

Fitting of the measured PPTR signals and DRS spectra with predictions of our numerical model (Sects. 2.2. and 2.3) constitutes a multi-dimensional optimization problem. The solution is sought by iterative minimization of the residual norm using a nonlinear least-squares algorithm (function *lsqnonlin* in Matlab; Mathworks, Natick, MA) (Fig. 2(a)). Specifically, the free parameters of our four-layer skin model are the epidermal and dermal thickness (d_{epi} and d_{der}), epidermal melanin and blood contents (m , b_{epi}), papillary and reticular dermal blood contents (b_{pap} and b_{der}), papillary and reticular blood oxygenation levels (S_{pap} and S_{ret}), scattering parameters of the epidermis and dermis (a_{epi} , p_{epi} , a_{der} , and p_{der}), and subcutis scattering amplitude (A). In addition to the listed skin properties we also optimize the radiant exposure applied in the respective PPTR measurement (F).

The initial 1.5 seconds of each PPTR signal (*i.e.*, 1500 uniformly distributed data points) are used in the analysis. Meanwhile, the DRS spectra are fitted at only 14 wavelengths, selected by considering the absorption spectra of skin chromophores (see Fig. 5), in order to reduce the computational load of the numerous MC runs.

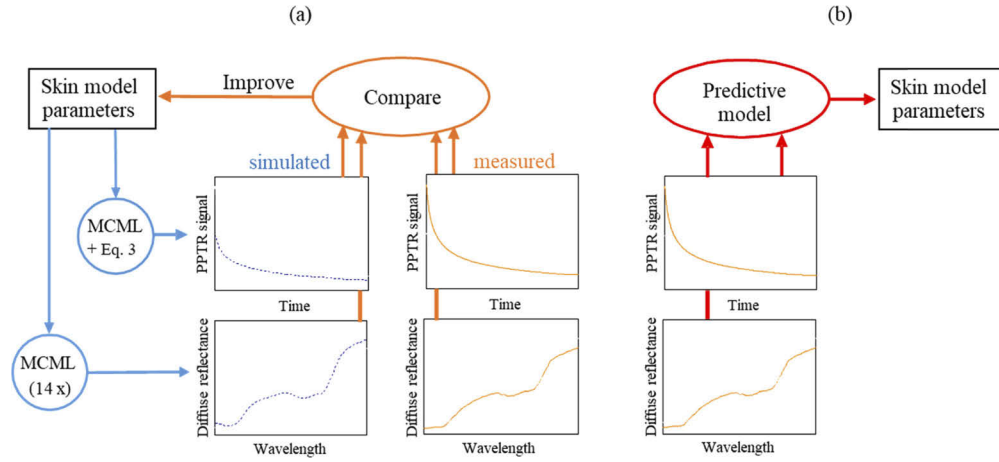


Fig. 2. Schematic representation of skin model parameter assessment using (a) iterative multi-dimensional optimization involving multiple Monte Carlo simulations in each loop (IMC), and (b) the predictive model.

For simultaneous fitting of both DRS and PPTR data we combine the respective residual norms into one cost function with a merging factor (M):

$$\varepsilon = \varepsilon_{\text{DRS}} + M\varepsilon_{\text{PPTR}}. \quad (4)$$

In the following we use $M = 10$, which was found earlier to be suitable for the problem at hand [1].

2.5. Preparation of the example data set

The data set used to construct our predictive model (PM) included 9029 sample “pairs”, consisting of different combinations of the above listed skin parameters and the corresponding PPTR signals and DRS spectra, computed using our numerical forward model.

In order to reduce the computational load, all PPTR signals in the sample data set were compressed from the original 1500 data points to a smaller subset of representative values. We achieve this by non-uniform (quadratic) signal binning, which draws from intrinsic properties of heat diffusion dynamics and performs well in the case of PPTR temperature depth profiling [23]. Specifically, the considered time points t_i scale with the square of uniformly distributed depths:

$$t_i = \frac{1}{2D} \left(\frac{z_{\max}}{N} i \right)^2, \quad (5)$$

where D marks skin’s thermal diffusivity, z_{\max} is a suitable depth (in our case 1 mm) and the integer i runs from 1 to the desired number of points (N).

Within the scope of this study we have constructed several PMs utilizing different compression levels ($N = 161, 82$, and 41) and also uncompressed PPTR signals. Because no notable differences in PM performance were observed, we limit this report in the following to the case of $N = 161$. This value represents a reasonable compromise between the computation load of the PM construction and the risk of information loss associated with stronger compression of the PPTR signals. (Additional information regarding this issue can be found in the Discussion.)

Similarly, the DRS spectra are represented by the values at the same 14 wavelengths as used in the IMC procedure.

2.6. Construction of the predictive model (PM)

Using the described sample data, we formulate the following multi-target regression problem: Given a set of 161 PPTR and 14 DRS features, predict the values of the 14 skin parameters that were used to generate them (Fig. 2(b)). We first decompose this task into 14 independent single-target regression problems, *i.e.*, predictions of one parameter value from the presented PPTR and DRS data. The solution of the multi-target regression problem is then obtained by concatenating the predictions from all single-target models into one prediction array.

Each single-target model is constructed using the random forest approach [10]. Each random forest model is an ensemble of 100 decision trees, constructed by combining bootstrap aggregation and feature subspace sampling. While bootstrap aggregation improves the stability of the classifier and reduces the variance of the predictions, feature subspace sampling reduces the correlations between the constituent trees by focusing them onto different feature subspaces.

The decision trees represent piece-wise constant functions, constructed by recursively splitting the feature space “greedily”, *i.e.*, by selecting the feature and the splitting point which maximizes the consequent reduction of variance in the target. Each tree is constructed by using a bootstrap sample (with replacements) of the original data, and by considering a subset of all 175 available features (randomly sampled) at each split. Finally, the predictions of all trees of the random forest are aggregated by averaging to yield the model prediction for the respective target parameter. Similarly, the standard deviation of the individual predictions enables estimation of the confidence interval of the aggregated prediction for a specific example.

The size of the considered feature subset was selected separately for each target from the set of $\{1, f^{1/2}, f/10, f/4, f/2, f\}$, where f marks the total number of available features ($f = 175$). The optimal choice was selected based on performance of the corresponding parameter-specific models as assessed by the cross-validation (CV) procedure.

2.7. Cross-validation testing (CV)

In each fold of the 10-fold CV, a separate PM is constructed using only 90% of the available examples (*a.k.a.* training set), and its predictive performance is assessed using the hold-out data (test set). For each parameter-specific model we calculate first the corresponding root-mean-squared error (RMSE):

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}, \quad (6)$$

where N is the number of examples in the test set, \hat{y}_i is the predicted value of the target parameter for the i -th example from the test set, and y_i is the true value of the target parameter for the same example.

In order to enable objective comparison of the performance for different parameters, which can have very different typical values (and measurement units), we compute also the corresponding coefficients of variation or *normalized RMSE*:

$$NRMSE = \frac{RMSE}{\langle y_T \rangle}, \quad (7)$$

where $\langle y_T \rangle$ represents the mean value of the target parameter within the training set.

In addition, because a low *NRMSE* is not always indicative of the actionable knowledge captured by the model, we assess also the so-called *relative RMSE*:

$$RRMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \langle y_T \rangle)^2}}. \quad (8)$$

This metric, customary in the field of machine learning, compares the predictive performance of the trained model against a “default” (naïve) model which always predicts the average value, $\langle y_T \rangle$. If the target values in the training and test sets belong to the same distribution, the *RMSE* of the default predictor is equal to the standard deviation of the target values. Consequently, if the PM is unable to capture more specific information than the simple average of each target parameter value, this will lead to $RRMSE \geq 1$. Conversely, the value of *RRMSE* would approach 0 in the case of ideal model performance.

Finally, we quantify the relationship between the predicted and true parameter values also by calculating the Pearson’s correlation coefficients for each single-target regression problem:

$$r = \frac{\sum_{i=1}^N (y_i - \langle y \rangle) (\hat{y}_i - \langle \hat{y} \rangle)}{\sqrt{\sum_{i=1}^N (y_i - \langle y \rangle)^2} \sqrt{\sum_{i=1}^N (\hat{y}_i - \langle \hat{y} \rangle)^2}}. \quad (9)$$

3. Results

3.1. Cross-validation testing of the predictive model (PM)

Figure 3 depicts the results obtained from all 10 folds of the CV process, which objectively estimates the performance of the constructed PM.

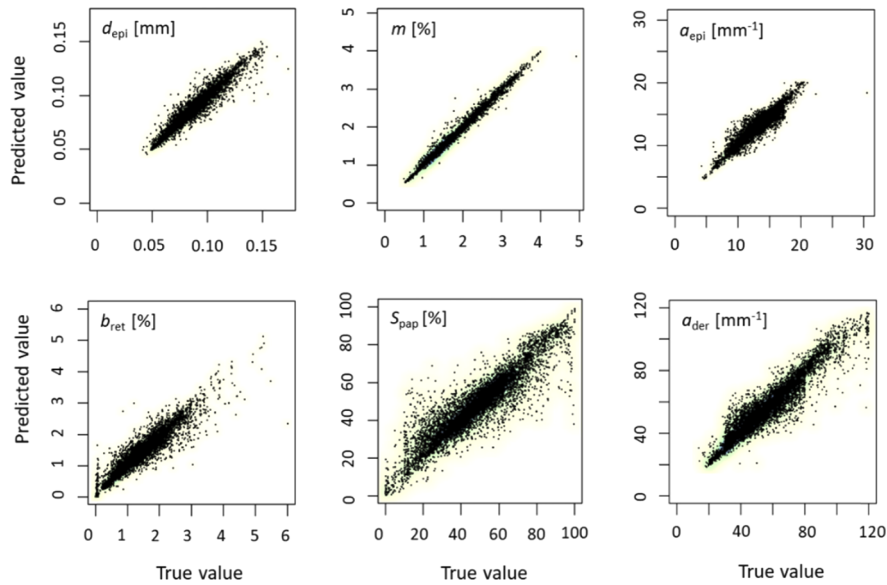


Fig. 3. Scatter plots of the predictive performance (cross-validation) for a subset of skin parameters. The deviations from the line of ideal performance ($\hat{y}_i = y_i$) indicate the absolute errors of the model predictions.

Note that we have initially constructed several random forests, using a different size of the feature subset at each split (see Sect. 2.6). In the following, we present only the results for the optimal size of the feature subset for each target parameter, as indicated by the lowest *RMSE* (Eq. (6)) obtained in the 10-fold CV. For most target parameters (specifically, 9 of 14) this means using the complete set of available features ($f = 175$) at each split. A feature subset size of $f/2$ was selected for three parameters (b_{ret} , a_{epi} , and p_{der}), and $f/4$ for the remaining two (b_{pap} and m), although the consequent improvement over the entire feature set was never significant.

Every point in these graphs represents one prediction of the selected target parameter value (\hat{y}_i) provided by the PM, constructed using the training set of the respective fold. For each point,

which comes from the respective test data set, the prediction is plotted against the true value, used in the forward MC model to compute the respective PPTR and DRS data (y_i). For an ideal PM, all points would thus lie on the diagonal line ($\hat{y}_i = y_i$).

The resulting values of the three performance metrics (Eqs. (6)–(8)) averaged across the 10 folds of the CV procedure are presented in Table 1. As evidenced by the lowest *NRMSE* values (in the range of 1–5%) the model predictions are most accurate for the applied radiant exposure (F), followed by the epidermal melanin content (m) and thickness (d_{epi}). Very good predictions are obtained also for both scattering amplitudes (a_{epi} and a_{der}) and the papillary blood content (b_{pap}), with *NRMSE* of 6–11%.

Table 1. Values of the PM performance metrics as obtained in CV testing. The values are averages across the 10 folds; for definitions of the metrics see Eqs. (6)–(9).

Parameter	<i>RMSE</i>	<i>NRMSE</i>	<i>RRMSE</i>	<i>r</i>
d_{epi}	0.005 mm	0.05	0.27	0.97
d_{der}	0.14 mm	0.19	0.40	0.92
m	0.06%	0.03	0.09	0.99
b_{epi}	0.10%	0.57	0.47	0.89
b_{pap}	0.32%	0.11	0.20	0.98
b_{ret}	0.22%	0.17	0.33	0.95
S_{pap}	7.1%	0.15	0.39	0.92
S_{ret}	11.1%	0.18	0.44	0.90
a_{epi}	0.8 mm^{-1}	0.06	0.33	0.98
a_{der}	0.6 mm^{-1}	0.10	0.30	0.93
p_{epi}	0.23	0.15	0.38	0.95
p_{der}	0.32	0.30	0.52	0.85
A	0.20	0.21	0.47	0.89
F	0.004 J/cm^2	0.01	0.16	0.99

The *RRMSE* values for the parameters listed above lie between 0.09 and 0.33, and the corresponding Pearson's correlation coefficients are very high ($r = 0.93$ – 0.99). This shows that the PM successfully captures a significant amount of information from the test examples.

The somewhat weaker performance observed for some other model parameters may be attributed in part to the specifics of the respective target value distributions in the available data set (see Fig. 4). Namely, the majority of examples were collected from the first 5–10 steps of the IMC analyses involving experimental data from human subjects. As a result, the respective target distributions reflect the natural distribution of anatomical and physiological properties within the involved population, rather than having a theoretically more advantageous uniform distribution. This is evident, *e.g.*, for the epidermal blood content (b_{epi} , Fig. 4(b)), which exhibits the worst performance in terms of *NRMSE*.

At the same time, however, the absolute values of b_{epi} are very small compared to both dermal blood contents (b_{pap} and b_{ret}). Consequently, the former are only weakly reflected in the features of both PPTR signals and DRS spectra by the nature of tissue optics, and are thus intrinsically difficult to assess by inverse analyses using any approach. In view of this, the corresponding Pearson's correlation coefficient is surprisingly high ($r = 0.89$).

3.2. Analysis of experimental data using the PM compared with IMC

Figure 5 presents the experimental PPTR signal and DRS spectrum acquired from the inner forearm of a healthy volunteer (27 years old man, JR) and the best-fitting MC results obtained

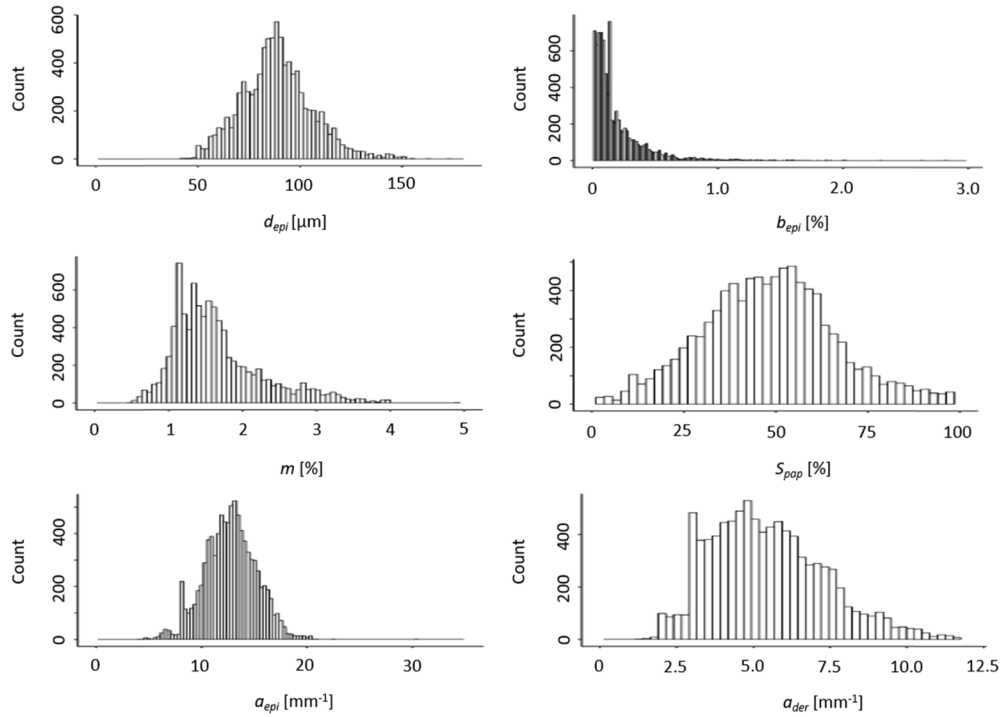


Fig. 4. Distribution of target values for selected model parameters within the training data set.

using our four-layer model of skin. The match between the measured data and predictions of our IMC model is excellent.

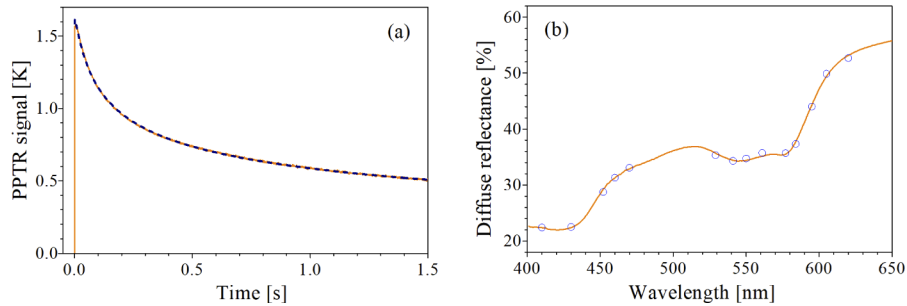


Fig. 5. PPTR signal (a) and DRS spectrum (b) as measured in volunteer JR (orange lines) and best fitting IMC results (blue dashed line and circles, respectively).

The obtained skin parameter values are listed in Table 2 (left column). All assessed values lie within anatomically and physiologically plausible ranges for healthy human skin [14,24–27]. This is in line with our earlier analyses of experimental PPTR and DRS data using the same approach [3,9,13,28].

Most importantly for the present study, the values predicted by the PM (middle column) lie mostly within the 95% confidence intervals (CI) of the IMC results. The latter were calculated as $CI = 2.78 SD / N^{1/2}$, where N marks the number of included IMC runs (in our case 5), and SD is the standard deviation of results from each run. For the PM, the CI amount to $1.98 SD / N^{1/2}$,

Table 2. The parameter values assessed by analysis of data from Fig. 5 using the IMC approach (left column), the predictive model (PM), and the hybrid approach (HA), with the corresponding 95% confidence intervals (CI) and their relative values (δ). The latter are rounded to 0 where smaller than 1 on the last displayed decimal place of the result.

Parameter	IMC	CI	δ	PM	CI	δ	HA	CI	δ
d_{epi} [mm]	0.106	0.009	0.08	0.104	0.002	0.02	0.124	0.007	0.06
d_{der} [mm]	0.93	0.16	0.17	0.87	0.04	0.05	1.14	0.14	0.12
m [%]	0.98	0.02	0.02	1.00	0.04	0.04	0.99	0.03	0.03
b_{epi} [%]	0.2	0.2	0.81	0.2	0	0.09	0.4	0.1	0.22
b_{pap} [%]	2.3	0.4	0.14	2.3	0.1	0.05	1.9	0.1	0.06
b_{ret} [%]	1.3	0.2	0.18	1.5	0.01	0.01	1.7	0.1	0.07
S_{pap} [%]	34	14	0.41	41	3	0.07	40	4	0.10
S_{ret} [%]	75	15	0.20	70	3	0.04	81	1	0.01
a_{epi} [mm ⁻¹]	11.1	1.9	0.17	10.7	0.3	0.03	10.7	0.2	0.02
a_{der} [mm ⁻¹]	4.0	0.6	0.15	4.2	0.1	0.02	5.6	0.2	0.05
p_{epi}	0.9	0.5	0.49	0.9	0.1	0.09	1.0	0.1	0.07
p_{der}	0.8	0.2	0.21	1.1	0.1	0.07	1.0	0.1	0.08
A	0.92	0.06	0.07	1.46	0.08	0.05	0.95	0.20	0.21
F [J/cm ²]	0.27	0	0	0.27	0	0	0.25	0	0
$\mathcal{E}_{\text{PPTR}}$	0.021	0.005		/	/		0.008	0.001	
\mathcal{E}_{DRS}	1.63	0.12		/	/		1.47	0.05	
\mathcal{E}	1.84	0.14		/	/		1.55	0.04	

where N is the number of decision trees (100). (The constants 2.78 and 1.98 are values of the t -distribution for $\alpha = 0.95$ with 4 and 99 degrees of freedom, respectively.)

Figure 6 presents a graphical comparison of the skin model parameter values obtained by using the IMC procedure and our PM from measurements in subject JR (data in Table 2). The values assessed using the two approaches evidently lie within the respective CI .

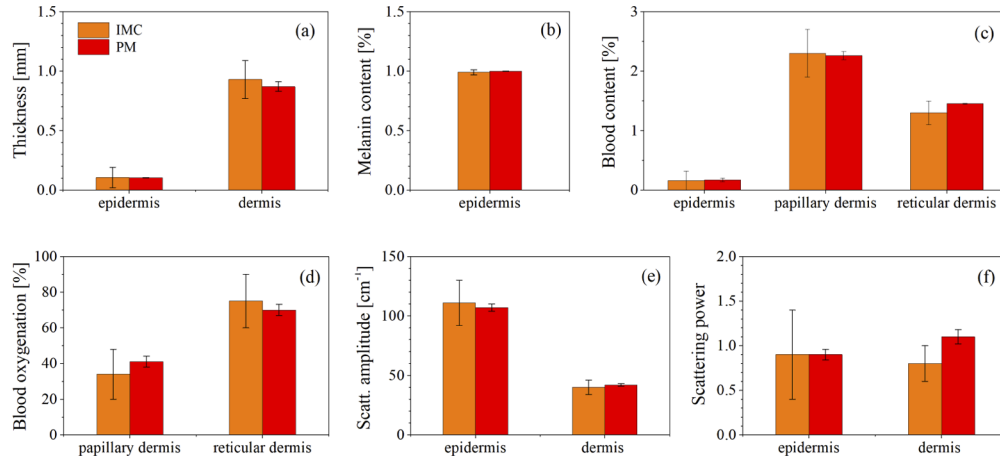


Fig. 6. Parameters of healthy human skin as assessed from measurements in subject JR using the IMC analysis (orange) and our PM (red). (data from Table 2)

We find this quite impressive, given the high dimensionality of our problem and the fact that the PM results are produced in just 0.15 s. In contrast, the IMC procedure typically involves

20–25 iteration steps, which take around 1 hour for a single run on a personal computer with 16 GB of RAM and a high-performance graphics card (NVidia GTX 1080).

In addition, the *CI* of all parameter values obtained from the PM are notably smaller than those from the IMC.

However, we can occasionally notice significant, albeit still relatively small differences between the IMC and PM results for some skin parameters. *E.g.*, in the analysis of data from another healthy volunteer (28 years old woman, NV), such deviations can be seen for the blood content and oxygenation in the reticular dermis (Figs. 7(c) and 7(d)), as well as for the scattering amplitude and power in the dermis (7(e) and 7(f)).

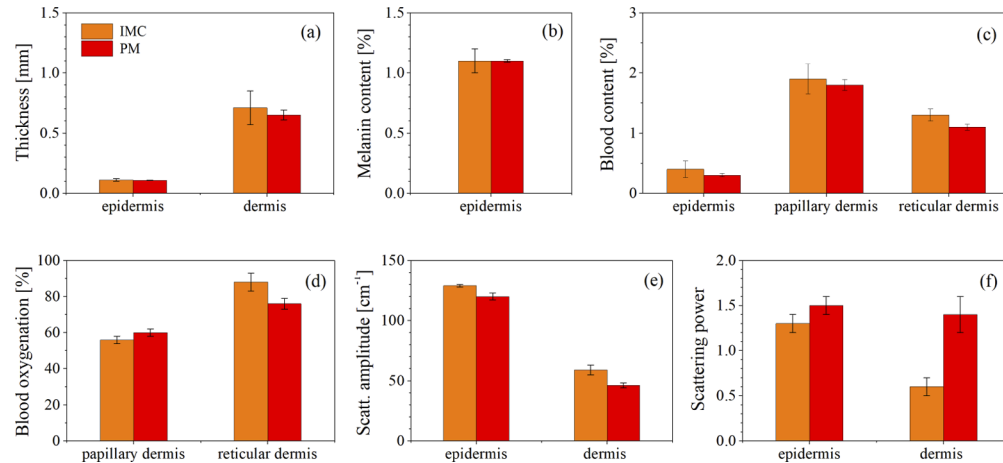


Fig. 7. Same as Fig. 6, but from measurements in another volunteer (NV).

In Fig. 8 we plot the spectral dependencies of the reduced scattering coefficients of the epidermis and dermis in both human subjects, as obtained from IMC analysis vs. the PM (see the legend). This is illustrative because they represent a collective effect of the model parameters a and p , which are independent variables in our inverse analyses.

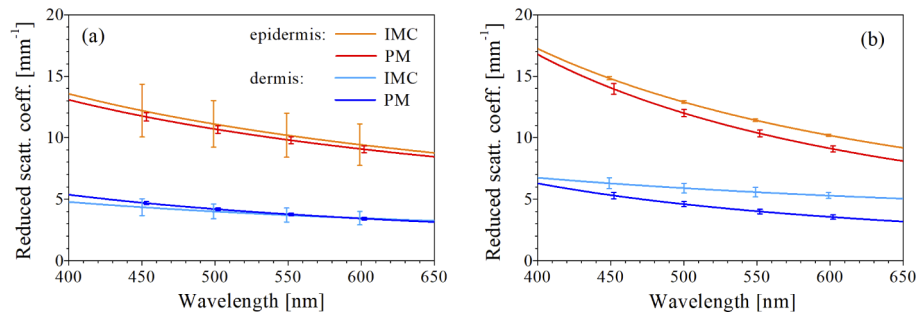


Fig. 8. Spectral dependence of reduced scattering coefficient of the epidermis and dermis in subjects JR (a) and NV (b), as obtained from IMC analyses and the PM (see the legend).

It is also instructive to compare experimental PPTR and DRS data with the MC predictions that correspond to the skin parameters assessed from the same using our PM. As shown in Fig. 9 for the case of subject JR, the match is reasonably good, but not as close as for the IMC result (Fig. 5). In the interest of conceptual clarity, only the 161 time points considered in construction of the PM are included in the presentation of the simulated PPTR signal (Fig. 8(a)).

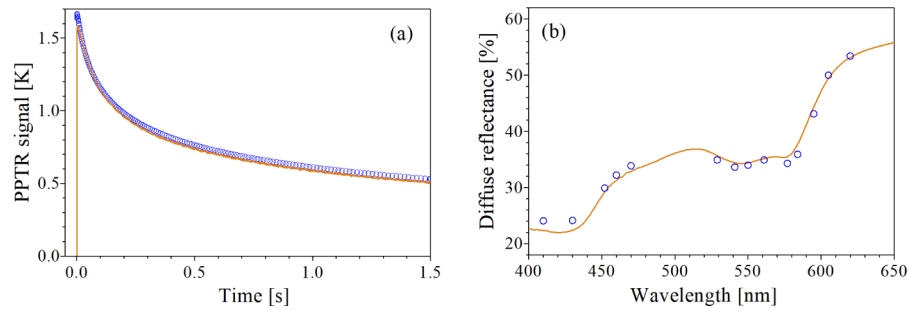


Fig. 9. PPTR signal (a) and DRS spectrum (b) as measured in subject JR (orange lines), compared with the forward MC predictions based on parameter values obtained with the PM (blue circles).

3.3. Hybrid approach

Based on the evidence presented in Figs. 6 and 7, we aim to further improve the accuracy and robustness of our inverse analysis. To that end, we discuss in the following a hybrid approach (HA), which combines the speed of PM with versatility of the iterative IMC procedure. Specifically, we use the PM result to initialize the IMC, which by default starts from randomly selected parameter values.

The skin parameter values assessed by using such an approach on data from subject JR are presented in Table 2 (right column). Again, all obtained values are anatomically and physiologically plausible for the involved test site and volunteers' age. Moreover, as illustrated in Fig. 10, the results obtained using the IMC, PM, and HA share several nontrivial properties. These include the higher blood content in the papillary compared to reticular dermis (Fig. 10(b)), which is consistent with the existence of an extensive capillary network in the papillary dermis. In addition, oxygen saturation in the papillary dermis is lower than in reticular dermis (Fig. 10(c)). The scattering amplitude in the epidermis is significantly higher compared with dermis (Fig. 10(d)) while the respective scattering powers are very similar (Fig. 10(e)) [19,29].

It is also evident that the variance of the values obtained with the HA (blue circles) tends to be smaller than for the original IMC procedure (orange). This holds especially for the dermal blood contents and oxygenation (Figs. 10(b) and (c)), as well as for the epidermal and dermal scattering amplitude and power (Figs. 10(d) and (e)). This is supported by the respective values of the *CI* in Table 2.

For some parameters, however, the average HA result differs from its IMC counterpart by more than would be expected based on the respective *CI*s (Table 2). Most notable examples in this regard are significantly higher values of b_{ret} ($1.7 \pm 0.1\%$) and a_{der} ($5.6 \pm 0.2 \text{ mm}^{-1}$) compared with the IMC results of $1.3 \pm 0.2\%$ and $4.0 \pm 0.6 \text{ mm}^{-1}$, respectively.

In order to analyze this effect, we compare in Fig. 11 the evolution of selected parameter values during the iterative IMC and HA runs. It is evident that with the latter approach (blue lines) the parameter values stabilize in just a few iteration steps, in contrast with >15 steps required with the customary IMC procedure. In addition, the final parameter values obtained with the HA are usually closest to the result from the IMC run which yields the lowest residual norm (indicated by solid orange lines).

Finally, Fig. 12 demonstrates that in all HA runs from Fig. 11, the residual norm ε (Eq. (4)) reaches lower values than in any of the 5 standard IMC runs. The resulting match with experimental data is thus even better than that achieved with the IMC (see Fig. 5), especially for the PPTR signal (see Table 2). This is not illustrated separately because the improvement over the already excellent match in Fig. 5 could not be perceived by naked eye.

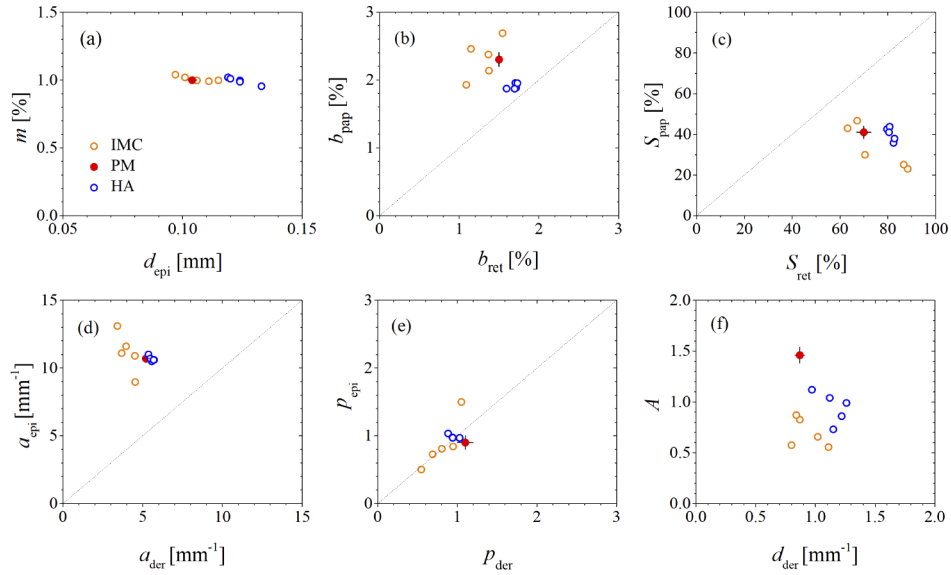


Fig. 10. Parameters of skin in subject JR as assessed using the IMC procedure (orange circles), our predictive model (red dots), and the hybrid approach (blue circles). The bars indicate the CIs of the corresponding PM results (see Table 2).

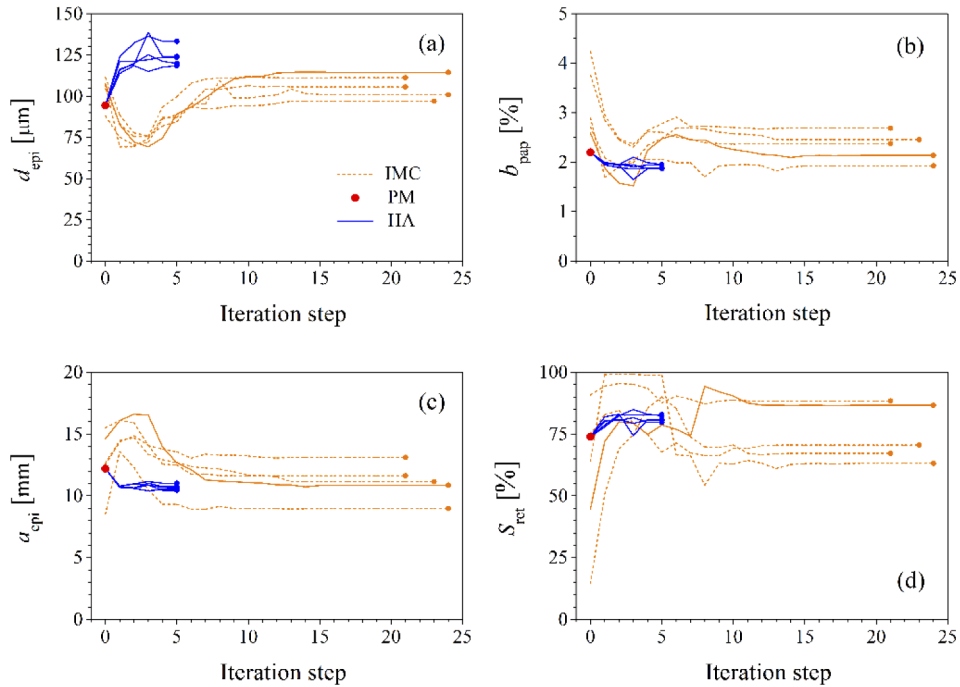


Fig. 11. Evolution of selected parameter values during five runs of the iterative IMC procedure with randomly selected initial values (orange lines), and with the initial values obtained using our PM (red dots) – *i.e.*, the hybrid approach (blue).

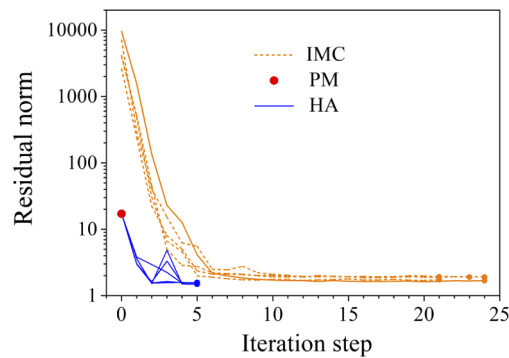


Fig. 12. Convergence of the IMC procedure with randomly selected initial values (orange lines) compared with the hybrid approach (blue), which is initialized by the PM result (red dot).

Moreover, the HA runs achieve this in 4–5 iteration steps, compared with the 20–25 steps required for convergence of the IMC procedure. Compared to the latter, the HA thus shortens the analysis time by a factor of ~ 5 , while at the same time providing a better match with experimental data and reduced variance of most skin parameter values.

4. Discussion

In this study, light transport and energy deposition in human skin are modeled using the Monte Carlo (MC) technique. This approach is known to provide realistic solutions of the radiative transport equation and is generally accepted as the “gold standard” in biomedical optics. Several versions of the MC technique coexist (*e.g.*, continuous vs. discrete absorption weighted), each with specific advantages and disadvantages, but they converge towards the same results in scattering dominated media with high scattering anisotropy [30,31]. We therefore use the practical and fast GPU-parallelized version of the MCML code by Alerstam *et al.* [22]

In the present work, we describe a machine learning approach to constructing a predictive model (PM) for quick quantitative assessment of multiple skin parameter values from measured PPTR signals and DRS spectra (Fig. 1). The involved multi-target regression problem is solved by using the random forests approach, which is robust and produces results with low bias. A collection of 9,029 examples from a forward numerical model of light-tissue interaction (Monte Carlo) served as the training data set used in construction of the 14 random forests, each of which contains 100 single-target decision trees (Sect. 2.6).

We show that the performance of such a PM as a surrogate for the computationally demanding IMC approach is very satisfying, both in objective testing using 10-fold cross-validation (Sect. 3.1) and in direct comparisons involving experimental data obtained from human volunteers (Sect. 3.2). Note that the latter tests involve experimental noise in the PPTR signals and DRS spectra (thus automatically with representative amplitudes and spectral distributions), as well as numerous intrinsic limitations and uncontrolled deficiencies of our 4-layer optical model (*e.g.*, the assumed chromophore spectra). Consequently, they represent a much more thorough and realistic test of the PM performance in real-life application, compared to what might be assessed using simulated data from the same forward MC model as used in training of PM with added synthetic noise.

In our experience thus far, predictions from the described PM are very close to the corresponding IMC results, most often within the respective 95% *CI* (*e.g.*, Table 2, Figs. 6 and 7). We find this quite impressive, given the high dimensionality of the problem (14 free variables) and relative sparsity of the training examples. In addition, the variance of parameter values assessed by using the PM is usually notably smaller in comparison with IMC (see the columns *CI* in Table 2).

Perhaps most importantly, the PM results are provided in just ~ 0.15 s, which is 10^5 times faster than our current IMC analyses. The latter namely take 5–6 hours on a high-performance personal computer, despite massive parallelization using the CUDA technology. In comparison, construction of each of the single-parametric random forests took only ~ 2 minutes, without any parallelization.

On the other hand, we can occasionally notice some differences between the PM and IMC-provided skin parameter values. However, these are most often limited to the properties which don't affect very strongly (and/or specifically) the measured PPTR signals and DRS spectra – *e.g.*, the dermal scattering amplitude and power (Figs. 7(e) and (f)). A more direct indication of the limited accuracy of our current version of the PM is thus the imperfect match between the input data and their numerically predicted counterparts based on the PM result (Fig. 8). A much closer match is evidently obtained with the IMC result (Fig. 5), although it is impossible to draw from this any quantitative conclusions about the errors of the individual skin parameter values constituting the PM result.

Before constructing the PM, all PPTR signals in the example data set were compressed from the original 1500 data points to 161 representative values by quadratic binning (Sect. 2.5) [23]. We therefore check whether the limited accuracy of our PM results might result from such a compression by analyzing separately its influence on the trusted IMC procedure.

In Fig. 13 we compare the experimental data from subject JR (orange lines) with the best fitting IMC predictions, obtained by using the quadratically binned PPTR signal (blue circles). The excellent match demonstrates that the discussed compression doesn't adversely affect the quality of the fit in comparison with the complete PPTR signal containing 1500 data points (Fig. 5). The DRS spectra are represented in both cases by the same 14 wavelengths.

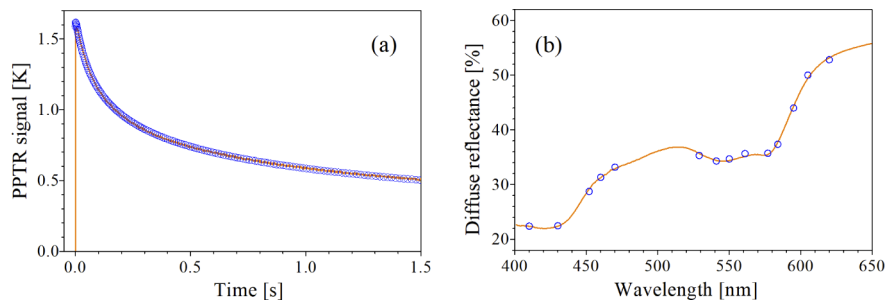


Fig. 13. PPTR signals (a) and DRS spectra (b) as measured in subject JR (orange lines) and the best fitting IMC results (blue circles) when compressing the PPTR signal to a subset of 161 data points by quadratic binning (Eq. (5)).

We have reached the same conclusion also in a similar analysis involving data from our second volunteer (NV). Moreover, even upon further signal compression (to 82 and 41 data points, respectively) the assessed parameter values and their variances didn't change significantly [28]. This proves that quadratic binning of the PPTR signals to 161 signal points did not contribute significantly to the limited accuracy of the PM results observed in the present study.

Overall, the presented PM captured very well the rather complex relationships between the 14 skin model parameters and the corresponding PPTR signals and DRS spectra (see Table 2). The information available directly from the training data set seemed sufficient for quantitative prediction of the relevant skin properties in the majority of analyzed examples. This was indicated also by our testing of PMs constructed using even more compressed PPTR signals ($N = 82$ and 41 ; not documented), which didn't indicate any improvement in terms of speed or performance. Application of more advanced feature construction methods (*e.g.*, an end-to-end approach using deep neural networks) was therefore not indicated for improvement of the performance.

More specifically, our provisional testing of the support vector regression (SVR) method with different kernels and neural networks (NN) indicated performance comparable to the presented PM. Moreover, performance of these machine learning models is very sensitive to the settings of several hyperparameters involved in their construction and training (*e.g.*, the choice of the kernel and its parameters, and parameter C for SVR; the number of hidden layers, the number of neurons in each layer, the learning rate, and the learning momentum for NN). Consequently, all listed hyperparameters must be carefully tuned for each intended application. This further increases the computational cost and complexity of these approaches and increases the risk of poor performance due to suboptimal settings of the hyperparameters.

On the other hand, we have also constructed PMs based on simpler generalized linear models and single decision trees. Their testing indicated unsatisfactory performance, which demanded application of a more powerful machine learning approach.

Based on the described arguments we have decided to apply random forests, which have a relatively small number of hyperparameters (*i.e.*, the number of trees and number of features selected at each split). Moreover, since random forests are robust with respect to the number of trees, the only hyperparameter that required tuning was the number of features selected at each split (see Sect. 3.1). In the discussed tuning process, we found that the *RMSE* values obtained with all feature subset sizes set to $f/2$ or $f/4$ were not significantly different from those achieved with the optimal model. This is in line with the default value of $f/3$ suggested by Breiman [10], and presents a viable option for simplified construction of PMs with insignificant loss of performance.

It is important to note that, primarily for practical reasons, most training examples used to construct our PM were obtained from IMC analyses of PPTR and DRS data from human subjects. Consequently, they don't sample uniformly the entire parameter space, but tend to reflect the natural distribution of anatomical and physiological properties within the involved population (see Fig. 4). We can therefore expect the PM to exhibit optimal performance in the most sampled part of the parameter space, but a limited performance for examples outside of this region. Further diversification of the training set might thus help to improve the performance of future PM.

Ideally, however, a PM deployed into a real-world setting would benefit from an incremental learning approach, such as learning on data streams. Such PMs don't need all the data up-front, but can update themselves as new data points arrive. Each new simulation is used first to test the current PM and then to update it [32]. Another machine learning methodology that could be considered for practical problems similar to the one discussed in our present study are the methods for multi-target regression. *E.g.*, a single ensemble of predictive clustering trees which predict all target parameters in parallel [33], rather than 14 ensembles of single-parametric decision trees as presented above. Ultimately, we can not exclude the use of more advanced feature construction methods, such as deep neural networks, especially if considering a more heterogeneous set of test sites and/or skin conditions.

For the time being, however, we have improved the accuracy of our inverse analysis of experimental data beyond that offered by the current PM, while still benefiting from its unparalleled speed, by using the hybrid approach (HA, Sect. 3.3). This approach involves using the PM result to initialize the iterative IMC procedure and provides a better match with experimental data than customary IMC and lower *CI*s for most skin parameter values (Table 2, Fig. 12). Moreover, this is achieved in just 5 iteration steps as opposed to the 20–25 steps required for convergence of the customary IMC (Figs. 11 and 12). The HA thus speeds up our data analysis by a considerable factor, while at the same time providing more accurate results with reduced variance.

Funding

Javna Agencija za Raziskovalno Dejavnost RS (N2-0128, P1-0192, P2-0103, PR-07590); Ministrstvo za Izobraževanje, Znanost in Šport (C3330-17-529021).

Acknowledgment

We thank Fotona d.o.o., Slovenia, for loan of the medical laser for PPTR measurements. Parts of this work were presented at BiOS 2019 [34] and ECBO 2019 [35].

Disclosures

The authors have no conflicts of interests to declare.

References

1. L. Vidovič, M. Milanič, L. L. Randeberg, and B. Majaron, "Quantitative characterization of traumatic bruises by combined pulsed photothermal radiometry and diffuse reflectance spectroscopy," *Proc. SPIE* **9303**, 930307 (2015).
2. N. Verdel, G. Lentsch, M. Balu, B. J. Tromberg, and B. Majaron, "Noninvasive assessment of skin structure by combined photothermal radiometry and optical spectroscopy: coregistration with multiphoton microscopy," *Appl. Opt.* **57**(18), D117–D122 (2018).
3. N. Verdel, A. Marin, M. Milanič, and B. Majaron, "Physiological and structural characterization of human skin in vivo using combined photothermal radiometry and diffuse reflectance spectroscopy," *Biomed. Opt. Express* **10**(2), 944–960 (2019).
4. T. E. Milner, D. M. Goodman, B. S. Tanenbaum, and J. S. Nelson, "Depth profiling of laser-heated chromophores in biological tissues by pulsed photothermal radiometry," *J. Opt. Soc. Am. A* **12**(7), 1479–1488 (1995).
5. L. Vidovič, M. Milanič, and B. Majaron, "Objective characterization of bruise evolution using photothermal depth profiling and Monte Carlo modeling," *J. Biomed. Opt.* **20**(1), 017001 (2015).
6. M. Milanič, I. Serša, and B. Majaron, "A spectrally composite reconstruction approach for improved resolution of pulsed photothermal temperature profiling in water-based samples," *Phys. Med. Biol.* **54**(9), 2829–2844 (2009).
7. A. Bashkatov, E. Genina, V. Kochubey, and V. Tuchin, "Optical properties of human skin, subcutaneous and mucous tissues in the wavelength range from 400 to 2000nm," *J. Phys. D: Appl. Phys.* **38**(15), 2543–2555 (2005).
8. P. Naglič, L. Vidovič, M. Milanič, L. L. Randeberg, and B. Majaron, "Suitability of diffusion approximation for an inverse analysis of diffuse reflectance spectra from human skin in vivo," *OSA Continuum* **2**(3), 905–922 (2019).
9. N. Verdel, A. Marin, L. Vidovič, M. Milanič, and B. Majaron, "In vivo characterization of structural and optical properties of human skin by combined photothermal radiometry and diffuse reflectance spectroscopy," *Proc. SPIE* **10037**, 100370H (2017).
10. L. Breiman, *Random Forests*, (Springer, 2001).
11. L. Vidovič and B. Majaron, "Elimination of single-beam substitution error in diffuse reflectance measurements using an integrating sphere," *J. Biomed. Opt.* **19**(2), 027006 (2014).
12. V. Dremine, E. Zherebtsov, A. Bykov, A. Popov, A. Doronin, and I. Meglinski, "Influence of blood pulsation on diagnostic volume in pulse oximetry and photoplethysmography measurements," *Appl. Opt.* **58**(34), 9398–9405 (2019).
13. A. Marin, N. Verdel, M. Milanič, and B. Majaron, "Influence of healthy skin baseline on bruise dynamics parameters as assessed by optical methods," *Proc. SPIE* **11075**, 110751O (2019).
14. S. L. Jacques, "Optical properties of biological tissues: a review," *Phys. Med. Biol.* **58**(11), R37–R61 (2013).
15. M. Friebe, A. Roggan, G. Müller, and M. Meinke, "Determination of optical properties of human blood in the spectral range 250 to 1100 nm using Monte Carlo simulations with hematocrit-dependent effective scattering phase functions," *J. Biomed. Opt.* **11**(3), 034021 (2006).
16. B. Choi, B. Majaron, and J. S. Nelson, "Computational model to evaluate port wine stain depth profiling using pulsed photothermal radiometry," *J. Biomed. Opt.* **9**(2), 299–307 (2004).
17. E.-J. Fiskerstrand, L. Svaasand, G. Kopstad, M. Dalaker, L. Norvang, and G. Volden, "Laser treatment of port wine stains: therapeutic outcome in relation to morphological parameters," *Br. J. Dermatol.* **134**(6), 1039–1043 (1996).
18. C. R. Simpson, M. Kohl, M. Essenpreis, and M. Cope, "Near-infrared optical properties of ex vivo human skin and subcutaneous tissues measured using the Monte Carlo inversion technique," *Phys. Med. Biol.* **43**(9), 2465–2478 (1998).
19. E. Salomatina, B. Jiang, J. Novak, and A. N. Yaroslavsky, "Optical properties of normal and cancerous human skin in the visible and near-infrared spectral range," *J. Biomed. Opt.* **11**(6), 064026 (2006).
20. A. Bashkatov, E. Genina, V. Kochubey, and V. Tuchin, "Optical properties of the subcutaneous adipose tissue in the spectral range 400–2500 nm," *Opt. Spectrosc.* **99**(5), 836–842 (2005).
21. L. Wang, S. L. Jacques, and L. Zheng, "MCML-Monte Carlo modeling of light transport in multi-layered tissues," *Comput. Meth Prog. Bio.* **47**(2), 131–146 (1995).
22. E. Alerstam, W. C. Y. Lo, T. D. Han, J. Rose, S. Andersson-Engels, and L. Lilge, "Next-generation acceleration and code optimization for light transport in turbid media using GPUs," *Biomed. Opt. Express* **1**(2), 658–675 (2010).
23. M. Milanič and B. Majaron, "Comparison of binning approaches for pulsed photothermal temperature profiling," *Proc. SPIE* **7371**, 73710O (2009).
24. R. I. Kelly, R. Pearse, R. H. Bull, J.-L. Leveque, J. de Rigal, and P. S. Mortimer, "The effects of aging on the cutaneous microvasculature," *J. Am. Acad. Dermatol.* **33**(5), 749–756 (1995).

25. D. Yudovsky and L. Pilon, "Retrieving skin properties from in vivo spectral reflectance measurements," *J. Biophoton.* **4**(5), 305–314 (2011).
26. C. Thorn, S. Matcher, I. Meglinski, and A. Shore, "Is mean blood saturation a useful marker of tissue oxygenation?" *Am. J. Physiol. Heart Circulat. Physiol.* **296**(5), H1289–H1295 (2009).
27. U. Merschbrock, J. Hoffmann, L. Caspary, J. Huber, U. Schmickaly, and D. Lübbbers, "Fast wavelength scanning reflectance spectrophotometer for noninvasive determination of hemoglobin oxygenation in human skin," *Int. J. Microcirc.* **14**(5), 274–281 (1994).
28. N. Verdel and B. Majaron, "Influence of signal binning on characterization of skin using photothermal radiometry and optical spectroscopy," Book of Abstracts, *The 20th International Conference on Photoacoustic and Photothermal Phenomena*, Moscow, Russia, 64–65 (2019).
29. W.-F. Cheong, "Summary of optical properties," *Optical-Thermal Response of Laser-Irradiated Tissue*, pp. 275–303 (Plenum, 1995).
30. C. K. Hayakawa, J. Spanier, and V. Venugopalan, "Comparative analysis of discrete and continuous absorption weighting estimators used in Monte Carlo simulations of radiative transport in turbid media," *J. Opt. Soc. Am. A* **31**(2), 301–311 (2014).
31. A. Doronin and I. Meglinski, "Online object oriented Monte Carlo computational tool for the needs of biomedical optics," *Biomed. Opt. Express* **2**(9), 2461–2469 (2011).
32. A. Osojnik, P. Panov, and S. Džeroski, "Tree-based methods for online multi-target regression," *J. Intell. Inf. Syst.* **50**(2), 315–339 (2018).
33. D. Kocov, C. Vens, J. Struyf, and S. Džeroski, "Tree ensembles for predicting structured outputs," *Pattern Recogn.* **46**(3), 817–833 (2013).
34. N. Verdel, J. Tanevski, S. Džeroski, and B. Majaron, "A machine-learning model for quantitative characterization of human skin using photothermal radiometry and diffuse reflectance spectroscopy," *Proc. SPIE* **10851**, 1085107 (2019).
35. N. Verdel, J. Tanevski, S. Džeroski, and B. Majaron, "Hybrid technique for characterization of human skin using a combined machine learning and inverse Monte Carlo approach," *Proc. SPIE* **11075**, 110751K (2019).